MODELING

Peer Review ETD
Experimental Toxicology Division
N H E E R L

A Multimodel Approach for Calculating Benchmark Dose

Ramon I. Garcia[1], R. Woodrow Setzer[2]

[1]Department of Biostatistics, School of Public Health, UNC [2] NHEERL/ETD/PKB

ENVIRONMENTAL ISSUE:

In the assessment of dose response, a number of plausible dose-response models may give fits that are consistent with the data. If no dose response formulation was specified *a priori* then the analyst is left with the decision of which model to select to represent the data. Typically, in this situation the analyst uses a model selection criterion to derive which model is superior to the others and proceeds to make inferences using this superior model. However, inferences made using the superior model will result in standard errors which are biased downward because they do not take model uncertainty into account. This effect can become significant when using the model to extrapolate away from the data *e.g.* estimation of the BMD (Benchmark Dose). Furthermore, it often transpires that more than one model adequately describes the data, and the best-fitting model is only trivially better than a number of others. In this case it becomes problematic to decide which of several models to use for a regulatory value. All of these issues fall under the general rubric “model uncertainty”.

RESEARCH GOALS:

The goals of this research are to explore and evaluate approaches to quantifying the uncertainty of dose-response predictions arising from uncertainty about the true dose-response model, and to develop practical implementations that can be incorporated in EPA’s BMDS software. The immediate goal is to evaluate the use of information theoretic model averaging for this purpose. The following interim goals define the overall progress of this project:

1. Develop computer code for commonly-used toxicological dose-response models for calculating the quantities required for information-theoretic model averaging.
2. For a selection of toxicological datasets, compare model-averaged estimates and confidence bounds to Bootstrap estimates that include model-selection as a characteristic. This determines whether the adjustments made by the information-theoretic methodology adequately account for model uncertainty in a realistic setting.
3. For a set of datasets simulated from “toy” mechanistic models (including PBPK and PD components), compare model-averaged BMDs and BMDLs to “true” BMD values derived from the models. This helps evaluate the extent to which information-theoretic confidence bounds are reasonable estimates of true BMDs derived under more-or-less realistic biological assumptions.
4. If warranted, develop a basis or draft of recommendations for the use of model-averaged estimates in regulatory dose-response assessment, including source-code for inclusion in EPA’s BMDS software package.

COMPUTATIONAL METHODS:

1. Model-Averaged Estimates: The information theoretic approach to model uncertainty is fully described in Burnham and Anderson (2002). In brief, model-averaged estimates for a benchmark dose for a dataset are calculated by:
   1. Fit (by maximum likelihood) plausible dose-response models to the dataset, and use each to calculate the BMDs of interest.
   2. Calculate Akaike Weights from the Akaike Information Coefficient for each model fit: $\exp(-AIC/2)$, normalized so that the weights sum to 1.0
   3. Inflate the widths of confidence intervals for each model to account for model selection.
   4. The model-averaged BMD and BMDL are the weighted average of the individual values.
2. Bootstrap:
   1. 1000 bootstrap samples are drawn from the original datasets
   2. All models are fit to each sample, and the best fitting model (the model with the lowest AIC) selected. In case of ties, one of the best-fitting models is selected at random.
   3. BMD estimates from the best-fitting models are accumulated over the 1000 bootstrap samples, and compared to the model averaged values for the original dataset.

CURRENT STATUS:

Code for the quantal models has been completed and is being tested for reliability and refined. Data sets for Goal 2 are being identified and collected.

NOTES on EXAMPLES

Table 1 and Figure 1:

- In this example, all models adequately describe the data, both from the standpoint of AIC and P-values, but BMD_{10}s have a nearly 2-fold range, and $BMDL_{10}$s a nearly 3-fold range.
- Bootstrap confidence intervals, including model selection, are 50% wider than the interval from BMD10 to BMDL10 calculated from best-fitting model, and the discordance between bootstrap, taken as close to truth and model based intervals increases for lower benchmark response rates (BMD_{01}, BMD_{001}).
- Model-averaged confidence limits more closely approximate the bootstrap confidence limits.
- Model averaging allows extrapolation to lower doses with a reasonable estimate of the reliability of those estimates.

Table 2 and Figure 2:

- Here, two models (probit and logistic) fail goodness of fit tests. However, the Akaike weights for them are so small that they contribute little to the overall average.
- The BMDs based on the best-fitting are biased, compared to bootstrap estimates. Model averaged estimates are less biased.

EXAMPLES

	Δ AIC[1]	Akaike Weights	GOF P-value[2]	$BMDL_{10}$	BMD_{10}	$BMDL_{01}$	BMD_{01}	$BMDL_{001}$	BMD_{001}
Log-Probit	0	0.317	0.75	80.11	102.19	28.18	35.95	13.13	16.75
Log-Logistic	1.47	0.152	1.0	46.31	90.94	4.21	17.85	0.42	3.72
Gamma	1.47	0.152	1.0	52.2	91.6	4.98	16.62	0.5	3.24
Weibul	1.47	0.152	1.0	52.2	92.05	4.98	15.32	0.5	2.64
Quantal Polynomial	1.47	0.152	1.0	52.2	93.66	4.98	10.97	0.5	1.12
Probit	4.01	0.043	0.19	118.77	148.21	18.26	29.99	1.96	3.55
Logistic	4.49	0.034	0.15	129.63	161.5	20.64	33.85	2.22	3.98
Model Averaged Estimates				61.28	99.99	3.82	23.03	0.16	7.22
Bootstrap Estimates				57.96	104.23	5.53	32.88		

[1]AIC of indicated model minus lowest AIC among suite of models

[2]Based on standard Chi-square goodness of fit test: values < 0.05 indicate significant lack of fit.

Table 1

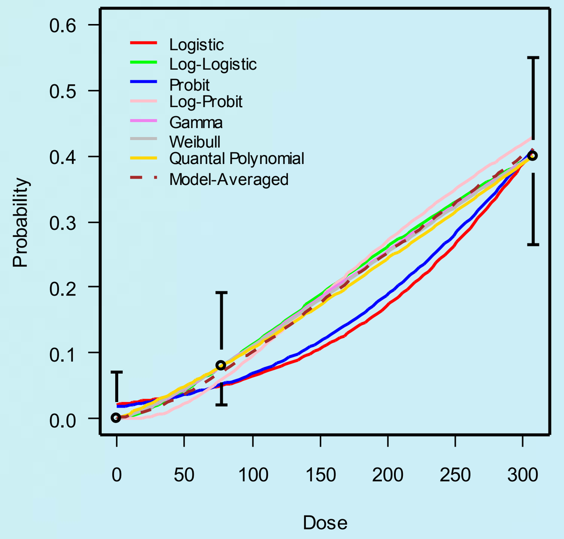

Figure 1

	Δ AIC	Akaike Weights	GOF P-value	$BMDL_{10}$	BMD_{10}	$BMDL_{01}$	BMD_{01}	$BMDL_{001}$	BMD_{001}
Gamma	0	0.259	0.99	9.46	11.92	0.9	1.14	0.09	0.11
Weibul	0	0.259	0.99	9.46	11.92	0.9	1.14	0.09	0.11
Quantal Polynomial	0	0.259	0.99	9.46	11.92	0.9	1.14	0.09	0.11
Log-Logistic	2	0.095	-[1]	4.34	27.11	0.39	7.79	0.04	2.34
Log-Probit	2	0.095	-	16.7	26.17	5.87	10.2	2.74	5.12
Logistic	5.2	0.019	0.02	25.84	33.17	2.92	4.06	0.3	0.42
Probit	5.82	0.014	0.02	25.86	32.2	2.85	3.79	0.29	0.39
Model Averaged Estimates				4.92	15.41	1.56	2.73	0.46	0.81
Bootstrap Estimates				6.06	23.15	0.59	6.06		

[1] residual df = 0

Table 2

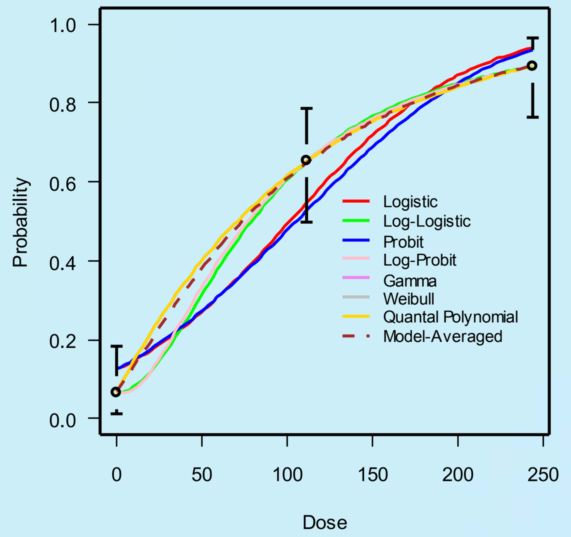

Figure 2

REFERENCES

Burnham KP and Anderson D. 2002. *Model Selection and Multi-Model Inference.* Springer. New York.

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

SOLVING AGENCY PROBLEMS